

Acquisition Research Program: Creating Synergy for Informed Change



LEXICAL LINK ANALYSIS (LLA) APPLICATION: IMPROVING WEB SERVICE TO DEFENSE ACQUISITION VISIBILITY ENVIRONMENT(DAVE)

May 13-14, 2015

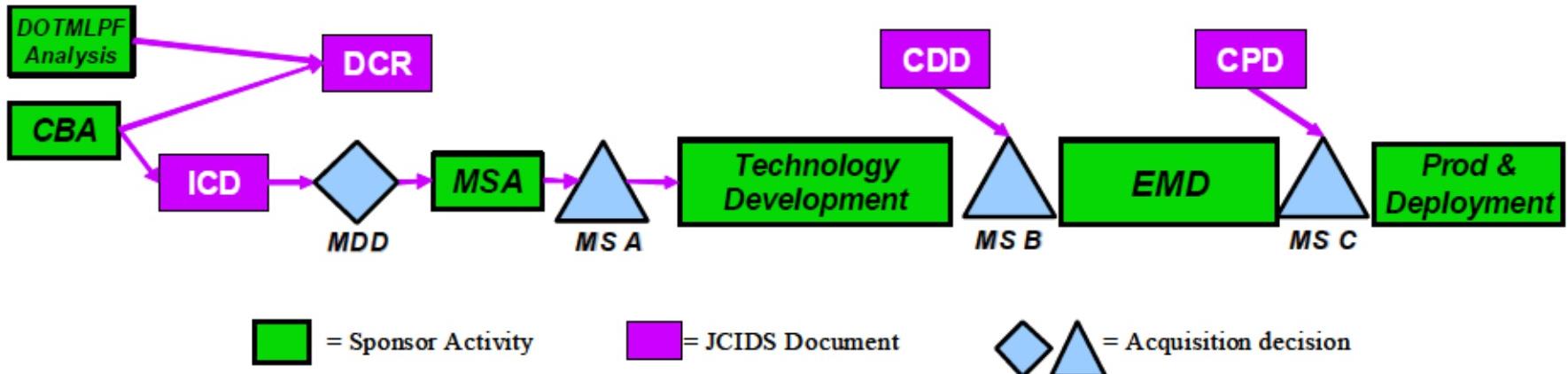
Dr. Ying Zhao, Dr. Douglas J. MacKinnon, Dr. Shelley P. Gallup,
Research Associate Professors

Distributed Information Systems Experimentation, Naval Postgraduate School

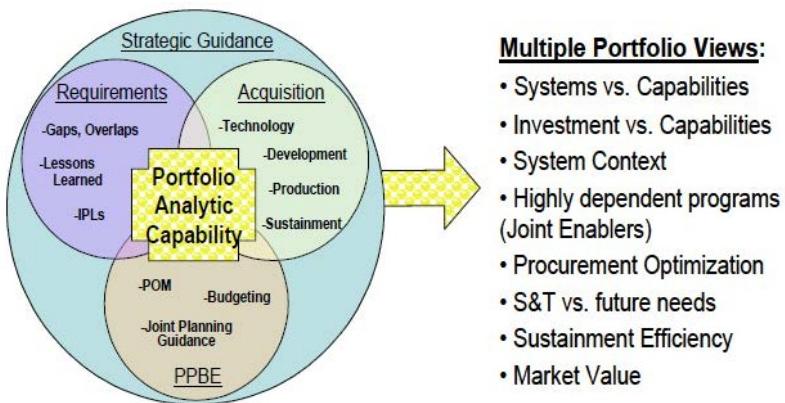
Report Documentation Page			Form Approved OMB No. 0704-0188	
<p>Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p>				
1. REPORT DATE MAY 2015	2. REPORT TYPE	3. DATES COVERED 00-00-2015 to 00-00-2015		
4. TITLE AND SUBTITLE Lexical Link Analysis (LLA) Application: Improving Web Service to Defense Acquisition Visibility Environment (DAVE)			5a. CONTRACT NUMBER	
			5b. GRANT NUMBER	
			5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)			5d. PROJECT NUMBER	
			5e. TASK NUMBER	
			5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School, Monterey, CA, 93943			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)	
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited				
13. SUPPLEMENTARY NOTES Presented at the 12th Annual Acquisition Research Symposium held May 13-14, 2015 in Monterey, CA.				
14. ABSTRACT				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 16
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified		



Background: Critical Needs of Automation, Validation, and Discovery



JCIDS Process and Acquisition Decisions (J-8 CJCSI 3170.01G)(JCIDS, 2009)

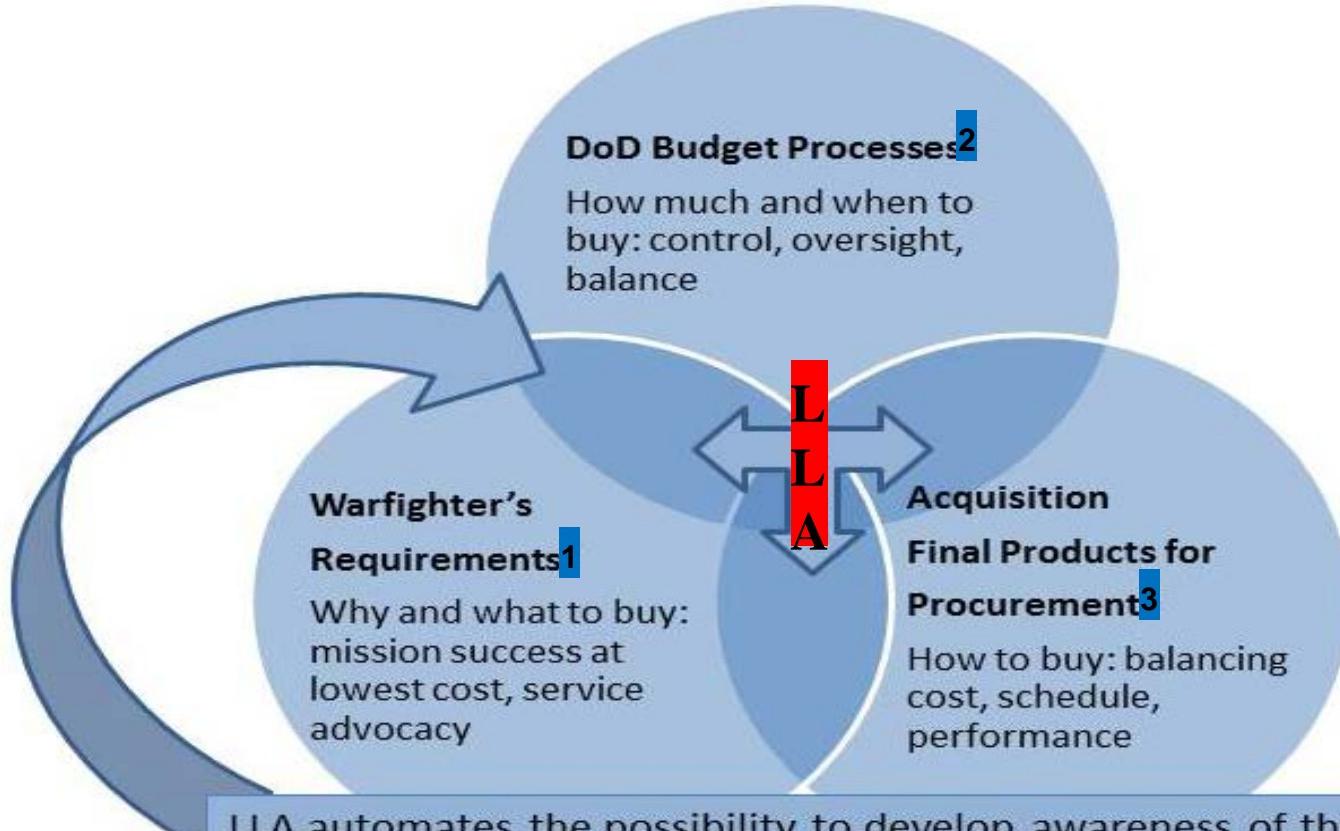


- Data are too voluminous, unformatted, and unstructured!
- Need to leverage automation
 - Extract relations among PE, MDAP, and ACATII
 - Extract costs





DoD Acquisition Decision Making



LLA automates the possibility to develop awareness of the "fit" between the budget, final products, and requirements.

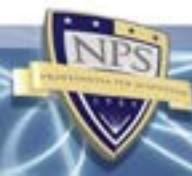
¹e.g. Universal Joint Tasks Lists, Urgent Needs Statements, ²e.g. Program Elements, ³e.g. Weapon Books





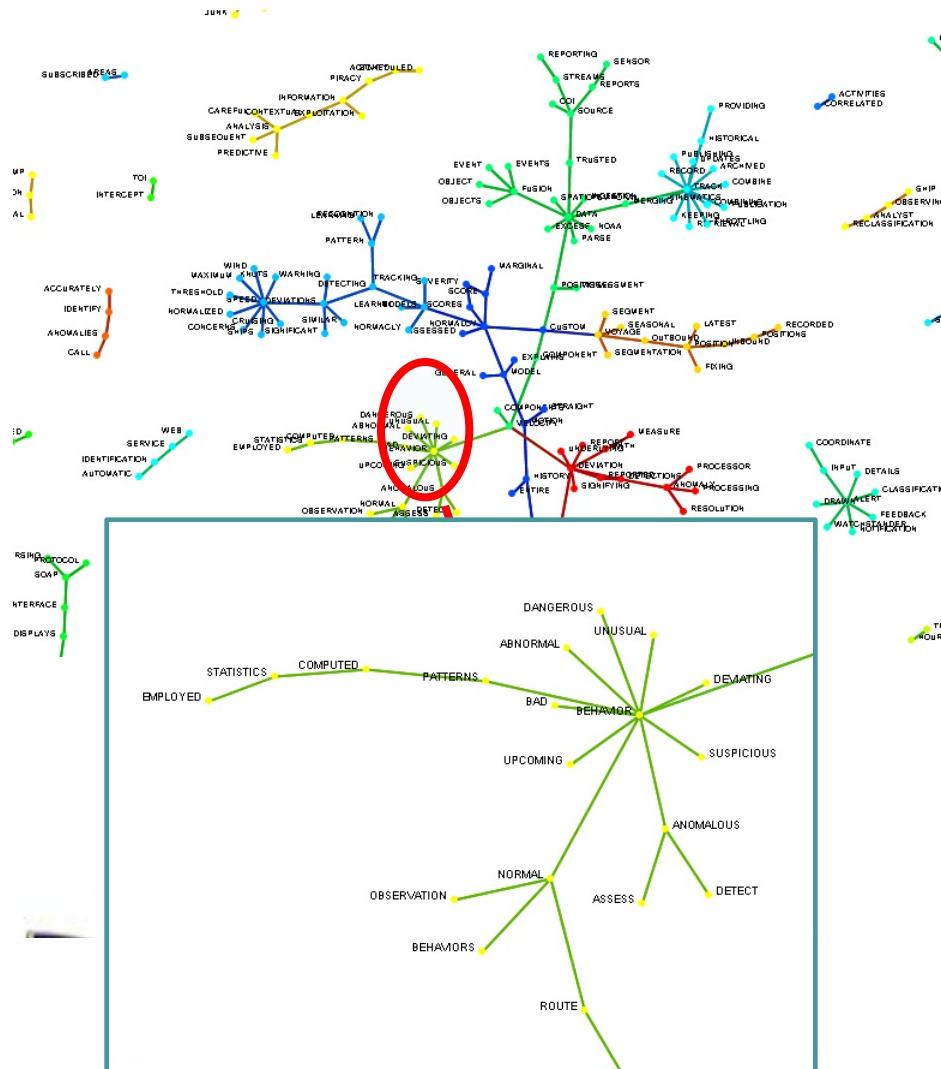
Methods

- Lexical Link Analysis (LLA) Core
 - LLA Reports and Visualizations
- Collaborative Learning Agents (CLA) for pattern recognition that scales up to Big Data
- System Self-Awareness (SSA)
- Big data and Deep Learning (BDDL) / Big Data Architecture and Analytics (BDDA)





Lexical Link Analysis (LLA) Core



Language-Independent Processing Steps

1. Extract word pairs (bigram) in context, connect them as word networks (text-as-networks)
 2. Group word networks into clusters /themes (by color)

Optional – Language-dependent Pre-processing Steps*

1. Stop-word Removal: Leave out stop words (e.g., a, of, the)
 2. Named Entity Extraction: Leave in/out people, places and organizations
 - Focus on semantics
 - Use NE in social networks
 3. Parts of Speech Tagging (Manning & Schütze, 1999): Use only nouns and verbs
 4. Stemming: Use only word roots

* Source: Stanford NLP (SNLP, 2012), Apache openNLP

Applying LLA as a Data-Driven Automation Technology and Methodology across the DoD Acquisition Process



- Surface themes and their relationships across multiple data sources
- Compare/correlate data from multiple data sources
- Sort/rank important and interesting information
- Discover high value areas for investment





Research Progress 2015

- **Task 1:** Work with the OUSD(AT&L) to install the LLA/CLA/SSA system as a web service in the Defense Acquisition Visibility Environment (DAVE) test bed via the AT&L eBusiness Center Service.
 - Approved and in progress.





Defense Acquisition Visibility Environment (DAVE)

- DAVE interfaces with authoritative data sources to support the OUSD(AT&L) and his staff in decision-making, oversight and analysis for major programs
- Sample data sources
 - **Defense Technical Information Center (DTIC)**
 - Program Elements (PEs):
[\[http://www.dtic.mil/descriptivesum/\]](http://www.dtic.mil/descriptivesum/)
 - **Defense Acquisition Management Information Retrieval (DAMIR)**
(<http://www.acq.osd.mil/dimir/>)
 - Selected Acquisition Report (SAR)
 - Defense Acquisition Executive Summary (DAES)
 - Acquisition Program Baseline (APB)
 - **Acquisition Information Repository (AIR)**
 - Test & Evaluation Master Plan (TEMP)
 - Systems Engineering Plan (SEP)
 - Life Cycle Sustainment Plan (LCSP)
 - Acquisition Strategy Reports (ASR)
 - Acquisition Decision Memorandum (ADM, for Milestone B 2366b Certification)
 - Technology Readiness Assessment (TRA), and LCSP





Example LLA Reports and Visualizations

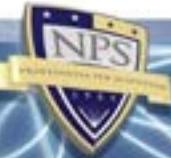
Event Date Sort|Theme Id|Theme Keywords

			Visualization	Matched	Unique	Total	Coupling	Gaps
10	70(E)	TOP 1000 HOT TOPICS IN BUSINESS	WordCloud	10	19	29	0.34	0.66
11	216(E)	TOP 1000 HOT TOPICS IN BUSINESS	WordCloud	10	19	29	0.34	0.66
12	85(E)	FINANCIAL CONDITION FINANCIAL WINNERS	WordCloud	10	19	29	0.34	0.66
13	47(E)	GLANCE COMPUTER FINAL GLANCE	WordCloud	18	35	53	0.34	0.66
14	11(E)	ANNOUNCEMENT SALE ANNOUNCES MONTHLY	WordCloud	15	36	45	0.33	0.67
15	84(E)	SHARES RECAP BUSINESS HOLD HOLDING SHARES	WordCloud	11	32	43	0.26	0.74
16	171(E)	BUSINESS RECAP BUSINESS RECAP BUSINESS WEEKLY BUSINESS RECAP	WordCloud	33	86	89	0.40	0.60
17	83(E)	BUSINESS RECAP BUSINESS RECAP BUSINESS WEEKLY BUSINESS RECAP	WordCloud	15	36	51	0.32	0.73
18	166(E)	WORLD MILITARY MEN	WordCloud	19	31	50	0.33	0.67
19	61(E)	HUT BUYS HUVS CHILDRENS	WordCloud	13	35	48	0.19	0.81
20	244(E)	HEALTH SERVICES HEALTH INFORMATION INFORMATION SERVICES	WordCloud	7	26	32	0.26	0.74
21	99(E)	CANADA GAS GAS PRICES	WordCloud	5	32	37	0.14	0.86
22	49(E)	SENIOR ANALYST ANALYST MEETING	WordCloud	13	46	53	0.25	0.75
23	36(E)	AIRLINE INDUSTRY AIRLINE CO2	WordCloud	10	31	41	0.24	0.76
24	236(E)	LEAD PLAINTIFF PERCUTANEOUS LEAD	WordCloud	8	32	40	0.20	0.80
25	18(E)	RANKSHARES CORP FILES SEC FORM RESOURCES CORP FILES SEC FORM	WordCloud	8	37	38	0.03	0.97
26	127(A)	PLANBAXY HOT HOT TRENDS	WordCloud	5	9	12	0.25	0.75
27	267(A)	NEUSTAR ANNOUNCED PACIFIC ANNOUNCED	WordCloud	8	29	37	0.22	0.78
28	87(A)		WordCloud					
29	101(A)		WordCloud					
30	215(A)		WordCloud					
31	70(A)		WordCloud					
32	37(A)		WordCloud					
33	245(A)		WordCloud					
34	91(A)		WordCloud					
35	120(A)		WordCloud					
36	176(A)		WordCloud					
37	111(A)		WordCloud					
38	148(A)		WordCloud					
39	5	Index_Technology	Match Score	252.00	—	121.00(0.26)	115.00(0.28)	123.00(0.36)
40	2	Index_Financial	Match Score	185.00	121.00(0.26)	—	97.00(0.25)	69.00(0.22)
41	3	Index_BasicMaterials	Match Score	176.00	115.00(0.28)	97.00(0.25)	—	84.00(0.29)
42	4	Index_ConsumerGoods	Match Score	154.00	123.00(0.36)	69.00(0.22)	84.00(0.29)	—
43	5	Index_Technology	Match Score	146.00	146.00(0.36)	93.00(0.25)	100.00(0.29)	87.00(0.31)
44	6	Index_Healthcare	Match Score	114.00	89.00(0.28)	67.00(0.23)	74.00(0.28)	68.00(0.31)
45	7	Index_IndustrialGoods	Match Score	89.00	70.00(0.25)	55.00(0.21)	57.00(0.25)	59.00(0.31)
46	8	Index_Utilities	Match Score	17.00	20.00(0.12)	22.00(0.15)	25.00(0.19)	17.00(0.15)
47			Match Score					
48			Index_Services					
49			Index_Financial					
50			Index_BasicMaterials					
51			Index_ConsumerGoods					
52			Index_Technology					
53			Index_Healthcare					
54			Index_IndustrialGoods					
55			Index_Utilities					
56			Index_Commodities					
57			Index_Demand					
58			Index_Supply					
59			Index_Production					
60			Index_Consumption					
61			Index_Exports					
62			Index_Imports					
63			Index_Production					
64			Index_Consumption					
65			Index_Exports					
66			Index_Imports					
67			Index_Demand					
68			Index_Supply					
69			Index_Production					
70			Index_Consumption					
71			Index_Exports					
72			Index_Imports					
73			Index_Production					
74			Index_Consumption					
75			Index_Exports					
76			Index_Imports					
77			Index_Demand					
78			Index_Supply					
79			Index_Production					
80			Index_Consumption					
81			Index_Exports					
82			Index_Imports					
83			Index_Production					
84			Index_Consumption					
85			Index_Exports					
86			Index_Imports					
87			Index_Demand					
88			Index_Supply					
89			Index_Production					
90			Index_Consumption					
91			Index_Exports					
92			Index_Imports					
93			Index_Production					
94			Index_Consumption					
95			Index_Exports					
96			Index_Imports					
97			Index_Demand					
98			Index_Supply					
99			Index_Production					
100			Index_Consumption					
101			Index_Exports					
102			Index_Imports					
103			Index_Production					
104			Index_Consumption					
105			Index_Exports					
106			Index_Imports					
107			Index_Demand					
108			Index_Supply					
109			Index_Production					
110			Index_Consumption					
111			Index_Exports					
112			Index_Imports					
113			Index_Production					
114			Index_Consumption					
115			Index_Exports					
116			Index_Imports					
117			Index_Demand					
118			Index_Supply					
119			Index_Production					
120			Index_Consumption					
121			Index_Exports					
122			Index_Imports					
123			Index_Production					
124			Index_Consumption					
125			Index_Exports					
126			Index_Imports					
127			Index_Demand					
128			Index_Supply					
129			Index_Production					
130			Index_Consumption					
131			Index_Exports					
132			Index_Imports					
133			Index_Production					
134			Index_Consumption					
135			Index_Exports					
136			Index_Imports					
137			Index_Demand					
138			Index_Supply					
139			Index_Production					
140			Index_Consumption					
141			Index_Exports					
142			Index_Imports					
143			Index_Production					
144			Index_Consumption					
145			Index_Exports					
146			Index_Imports					
147			Index_Demand					
148			Index_Supply					
149			Index_Production					
150			Index_Consumption					
151			Index_Exports					
152			Index_Imports					
153			Index_Production					
154			Index_Consumption					
155			Index_Exports					
156			Index_Imports					
157			Index_Demand					
158			Index_Supply					
159			Index_Production					
160			Index_Consumption					
161			Index_Exports					
162			Index_Imports					
163			Index_Production					
164			Index_Consumption					
165			Index_Exports					
166			Index_Imports					
167			Index_Demand					
168			Index_Supply					
169			Index_Production					
170			Index_Consumption					
171			Index_Exports					
172			Index_Imports					
173			Index_Production					
174			Index_Consumption					
175			Index_Exports					
176			Index_Imports					
177			Index_Demand					
178			Index_Supply					
179			Index_Production					
180			Index_Consumption					
181			Index_Exports					
182			Index_Imports					
183			Index_Production					
184			Index_Consumption					
185			Index_Exports					
186			Index_Imports					
187			Index_Demand					
188			Index_Supply					
189			Index_Production					
190			Index_Consumption					
191			Index_Exports					
192			Index_Imports					
193			Index_Production					
194			Index_Consumption					
195			Index_Exports					
196			Index_Imports					
197			Index_Demand					
198			Index_Supply					
199			Index_Production					
200			Index_Consumption					
201			Index_Exports					
202			Index_Imports					
203			Index_Production					
204			Index_Consumption					
205			Index_Exports					
206			Index_Imports					
207			Index_Demand					
208			Index_Supply					
209			Index_Production					
210			Index_Consumption					
211			Index_Exports					
212			Index_Imports					
213			Index_Production					
214			Index_Consumption					
215			Index_Exports					
216			Index_Imports					
217			Index_Demand					
218			Index_Supply					
219			Index_Production					
220			Index_Consumption					
221			Index_Exports					
222			Index_Imports					
223			Index_Production					
224			Index_Consumption					
225			Index_Exports					
226			Index_Imports					
227			Index_Demand					
228			Index_Supply					
229			Index_Production					
230			Index_Consumption					
231			Index_Exports					
232			Index_Imports					
233			Index_Production					
234			Index_Consumption					



Big Data and Deep Learning (BDDL)

- **Task 2:** We are also exploring how to use LLA jointly with other business intelligence tools especially Big Data Architecture and Analytics (BDAA) tools





- Data Storage
 - Safe and fault-tolerant
 - SQL vs. NoSQL
 - Graph databases
 - Unstructured data
 - Cloud infrastructures
- Operational Systems
 - real-time
 - no- or low-latency of response
- Analytics
 - Map/Reduce paradigm related to data sciences
 - Map: parallel computing
 - Reduce: data fusion
 - Apache Mahout , Spark

New Tools

- Challenge
 - Traditional data sciences algorithms
 - Tightly coupled Map/Reduce computations
 - Need to augment traditional data sciences with the “Map/Reduce” paradigm to take advantage of massive distributed parallel computing

Big Data and Deep Learning

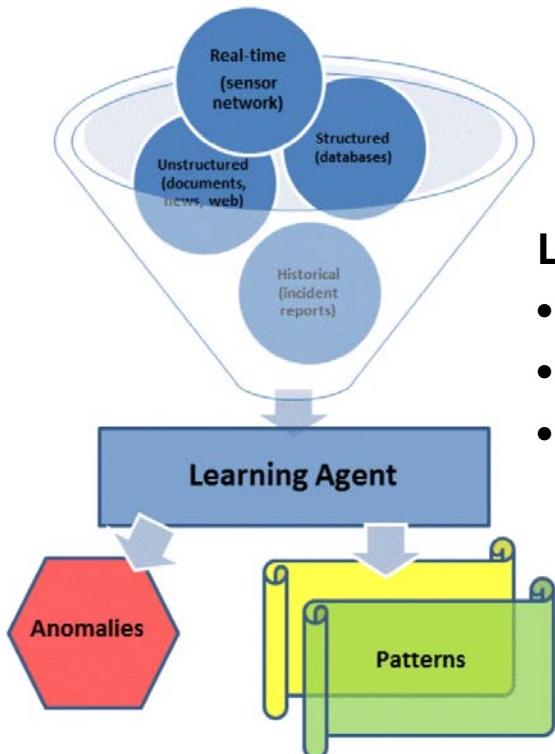
Commercial Trends for Analytics

- Predictive analytics
 - Identify high-valued targets
- Graph analytics
 - Core driver for Big Data analytics
- Text analysis
 - Data-driven discovery of themes and concepts
- Recommendation engine
 - Recommend high-valued targets explicitly
- Deep learning (especially useful for CID)
 - Self-taught learning: e.g. discover/identify interesting objects (e.g. cats and dogs) in massive images and videos
<http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html>
 - Unsupervised machine learning
 - Pattern recognition
 - Anomaly detection
 - Data fusion
 - Abundant DoD applications

Zhao, Y., MacKinnon, D.J., and Gallup, S.P. (August, 2015). Big Data and Deep Learning for Understanding DoD Data. In the CrossTalk, the Journal of Defense Software Engineering, Special Issue “Data Mining and Metrics.”

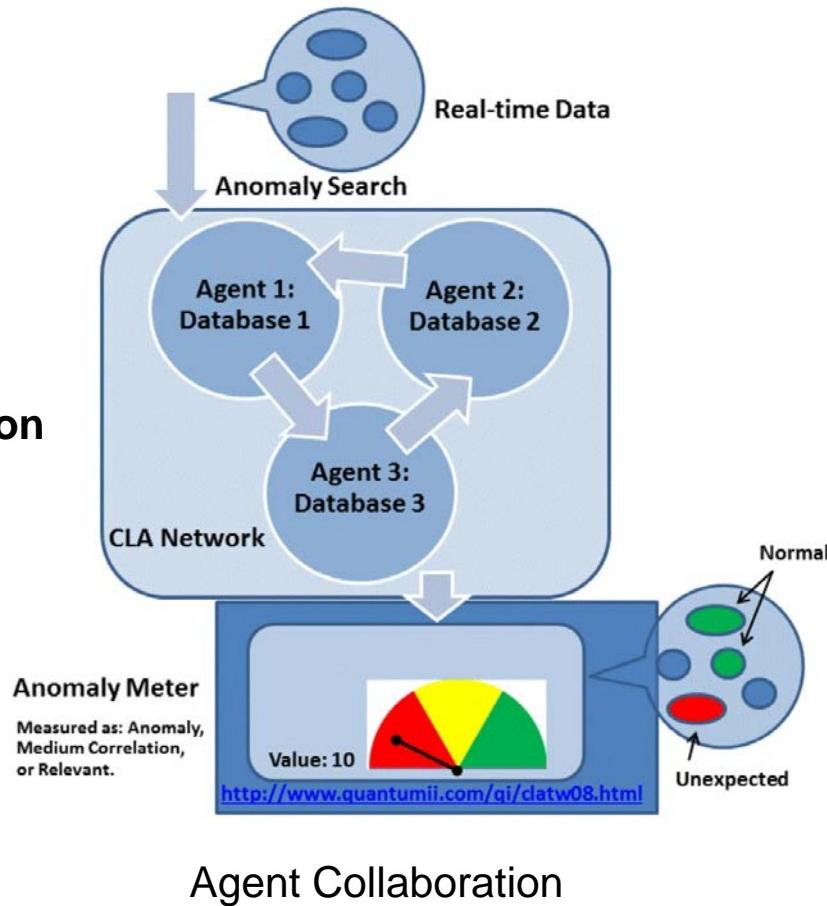


Collaborative Learning Agent (CLA): It is Big Data and Deep Learning!



A Single Collaborative Learning Agent (CLA) Patterns are graded from medium to relevant correlations.

- LLA is Deep Learning**
- Pattern Recognition
 - Anomaly Detection
 - Recursive Data Fusion



Agent Collaboration

System and Method for Knowledge Pattern Search from Networked Agents (US patent 8,903,756)



How Big Data and Deep Learning can be Applied to the DoD Acquisition Process

- Examine and compare the important acquisition data sources to gain business insights:
 - Study Prime and subcontractor relationships
 - DOT&E Annual Reports for horizontal analysis
 - Compare DOT&E and DAES: What did the OT result say and did issues surface?
 - Did these issues show up somewhere in DAES earlier?
 - Time series analysis
 - Can LLA look at program ratings over time?
 - Compare budget data and contracts data





How Big Data and Deep Learning can be Applied to the DoD Acquisition Process

(cont'd)

- In the current acquisition process, a small delay or anomaly in a contract negotiation process can have a huge impact in its performance, therefore cost the government a lot of money downstream.
- It will be very useful to apply BDAA such as LLA for pattern recognition and anomaly detection for these kind of problems and make early warnings and predictions to prevent the downstream risks.
- The Big Acquisition Data might include programs' cost, SAR, DAMIR, tech data, even outside economic environment data if the access is possible.
- The causes of the deviations from the normal behaviors for the programs/contracts might be modeled using physics (e.g., fluid dynamics theories).





How Big Data and Deep Learning can be Applied to the DoD Acquisition Process

(cont'd)

- LLA's network perspectives, social plays among the nodes and the System Self-Awareness (SSA) theory may be used to lay out the academic vigor for the business processes, for example, answer the questions
 - Are some nodes drawn towards some other nodes because the other nodes are more powerful?
 - Is the preferential attachment growth pattern or expertise growth pattern can be used here?
 - How are the forces of the nodes modeled and mapped into the social network settings and actual business processes?





Acknowledgements

- We thank Mr. Robert Flowe from OUSD(AT&L)/ARA who provided sponsorship to access the Defense Acquisition Visibility Environment, and relevant questions along with insightful guidance and discussions.

